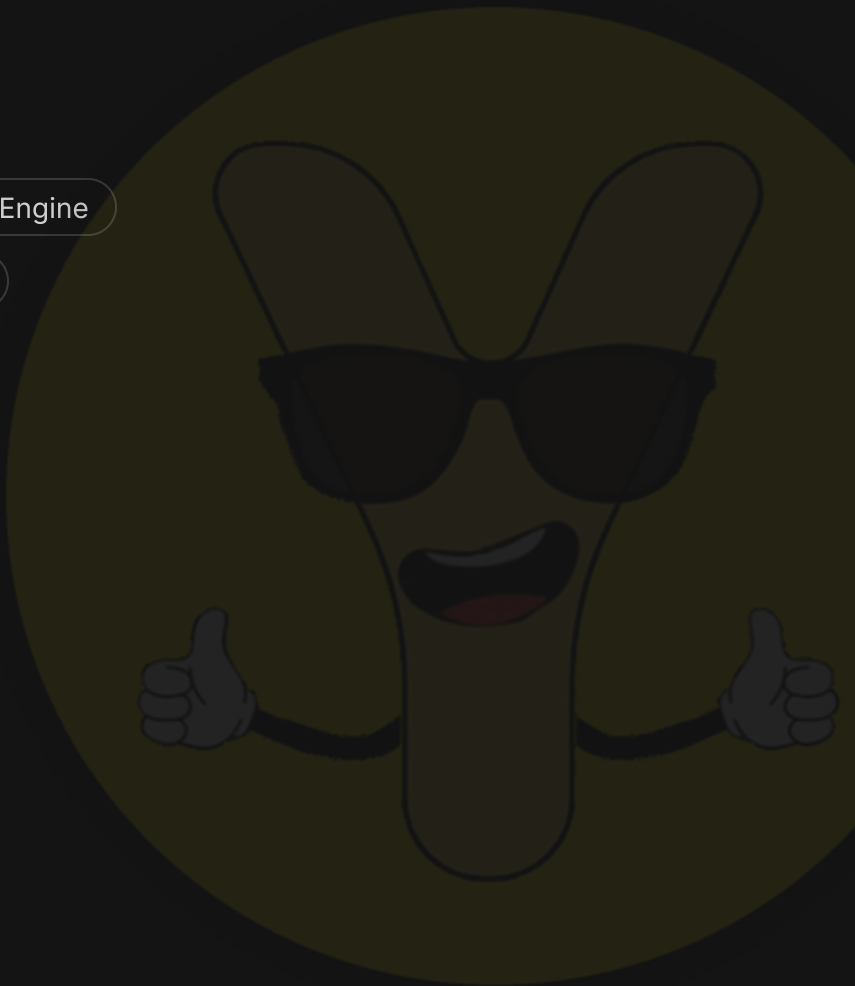
AKTU · 4TH YEAR · SEMESTER 7 · BCS071

Cloud Computing

Unit 5 — Cloud Technologies And Advancements

- Hadoop
- MapReduce
- Virtual Box
- Google App Engine
- Programming Environment for Google App Engine
- Open Stack
- Federation in the Cloud
- Four Levels of Federation
- Federated Services and Applications
- Future of Federation



DECODED BY

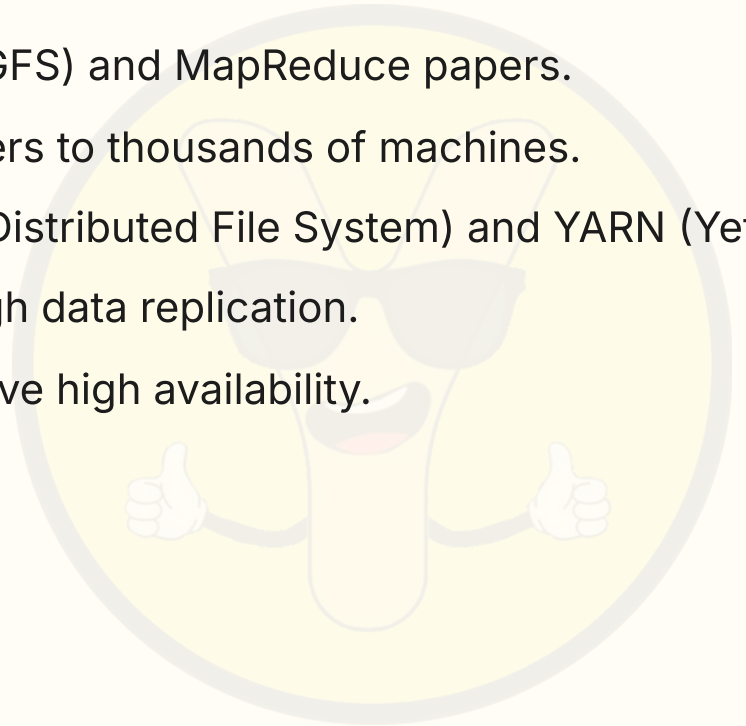
Chirag Jain

Founder, Ycotes · Gold Medalist in Artificial Intelligence, IIT Madras · taught 10,000+ students

Introduction to Hadoop

Hadoop is an open-source framework for distributed storage and processing of large datasets.

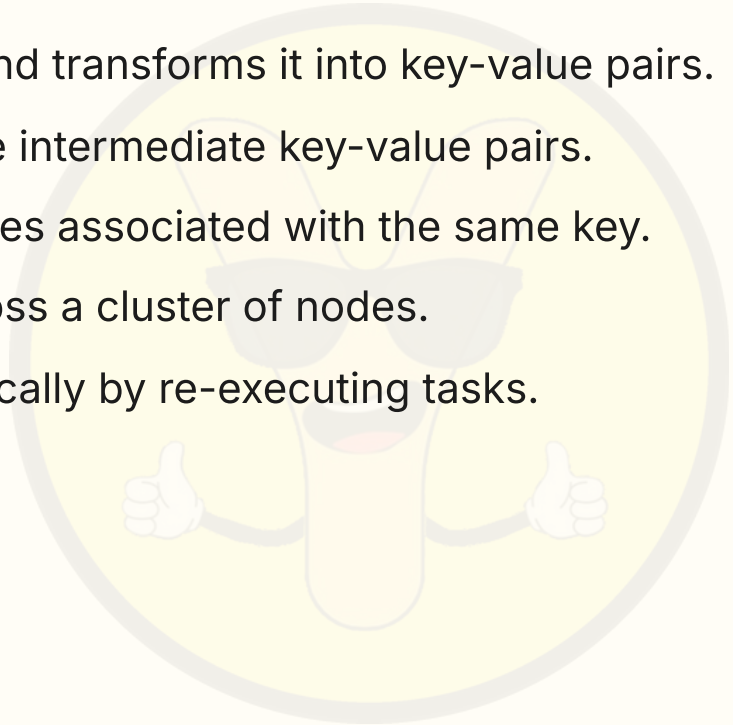
- Based on the Google File System (GFS) and MapReduce papers.
- Designed to scale from single servers to thousands of machines.
- Core components: HDFS (Hadoop Distributed File System) and YARN (Yet Another Resource Negotiator).
- Provides high fault tolerance through data replication.
- Uses commodity hardware to achieve high availability.



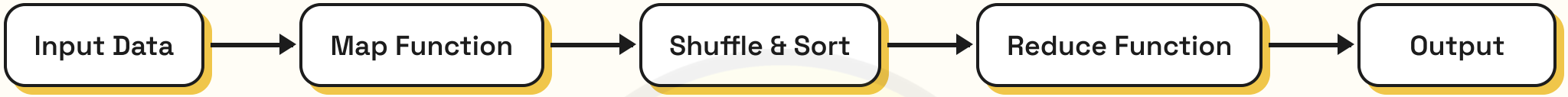
MapReduce Paradigm

MapReduce is a programming model for processing massive data sets in parallel.

- Map Phase: Processes input data and transforms it into key-value pairs.
- Shuffle Phase: Sorts and groups the intermediate key-value pairs.
- Reduce Phase: Aggregates the values associated with the same key.
- Enables distributed processing across a cluster of nodes.
- Handles hardware failures automatically by re-executing tasks.



MapReduce Workflow



The sequential flow of data processing in MapReduce.

VirtualBox

VIRTUALBOX

A cross-platform virtualization software that allows running multiple operating systems simultaneously on a single host machine.

- Type 2 Hypervisor: Runs on top of an existing host OS.
- Supports guest OS including Windows, Linux, Solaris, and BSD.
- Provides Virtual Machine (VM) snapshots for state recovery.
- Uses Guest Additions to improve performance and integration.

Google App Engine (GAE)

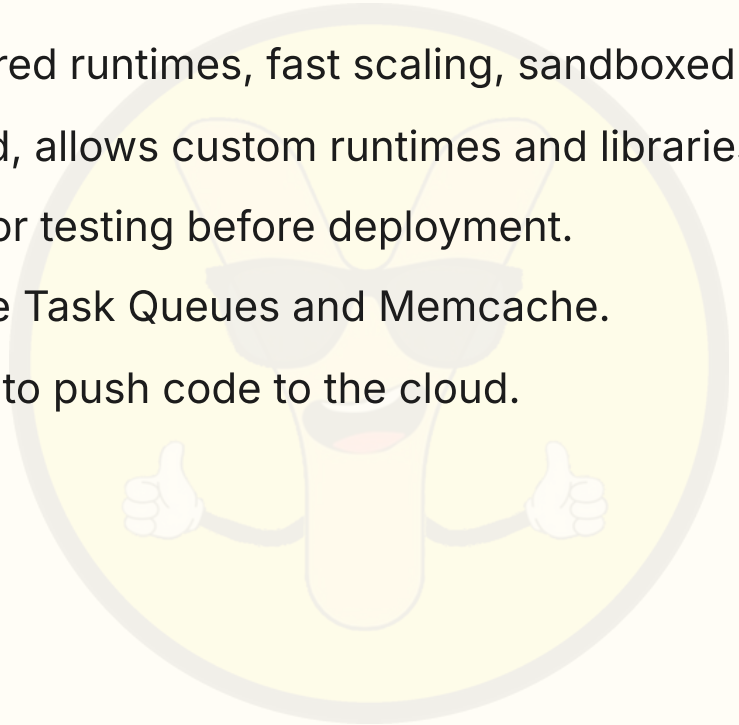
GAE is a Platform-as-a-Service (PaaS) offering by Google for building and hosting web applications.

- Fully managed environment: Google handles infrastructure, patching, and scaling.
- Automatic scaling: Resources scale up or down based on traffic.
- Supports multiple languages: Java, Python, PHP, Go, Node.js.
- Integrated with Google Cloud services like Datastore and Cloud SQL.
- Pay-as-you-go pricing model based on resource consumption.

Programming Environment for GAE

GAE provides specific environments to run applications effectively.

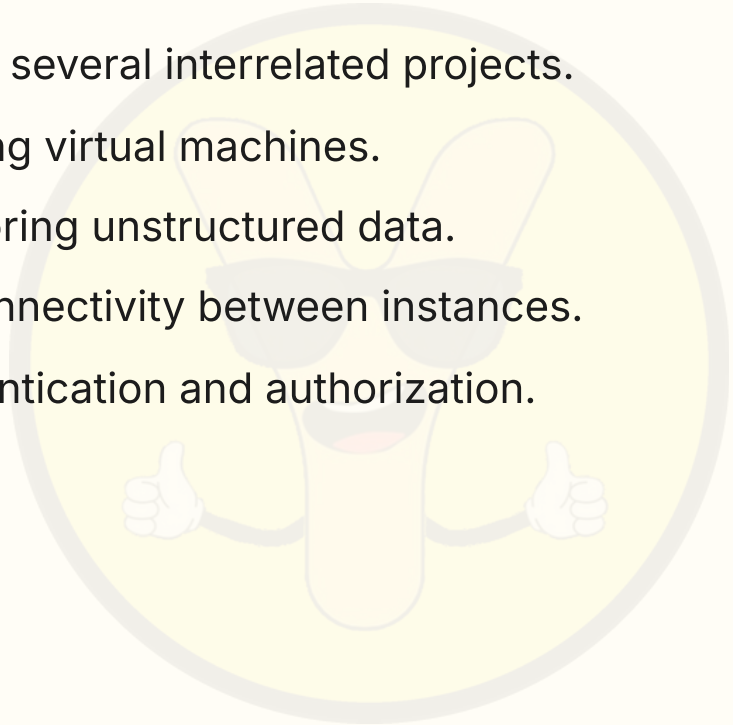
- Standard Environment: Pre-configured runtimes, fast scaling, sandboxed execution.
- Flexible Environment: Docker-based, allows custom runtimes and libraries.
- SDKs: Local development servers for testing before deployment.
- APIs: Access to built-in services like Task Queues and Memcache.
- Deployment: Use 'gcloud' CLI tools to push code to the cloud.



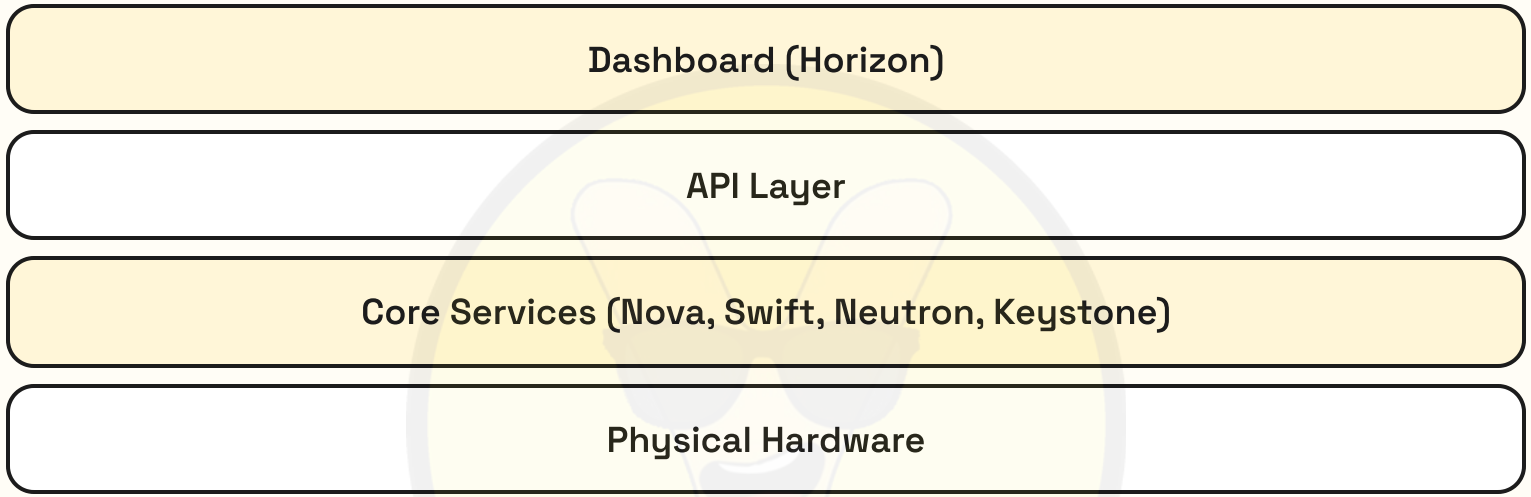
OpenStack

OpenStack is an open-source cloud computing platform for building private and public clouds.

- Modular architecture: Composed of several interrelated projects.
- Nova: Compute service for managing virtual machines.
- Swift: Object storage system for storing unstructured data.
- Neutron: Networking service for connectivity between instances.
- Keystone: Identity service for authentication and authorization.



OpenStack Architecture



Layered representation of OpenStack service components.

Federation in the Cloud

CLOUD FEDERATION

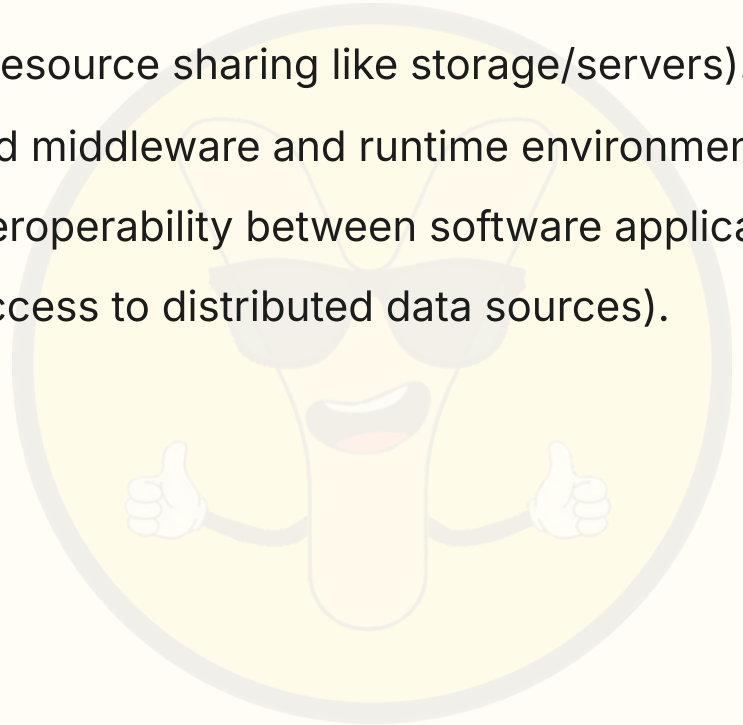
The practice of interconnecting cloud computing environments from different providers to allow interoperability and resource sharing.

- Allows a cloud provider to extend its capacity by using resources from another provider.
- Enables seamless migration of workloads between different clouds.
- Reduces vendor lock-in by providing a unified interface.
- Facilitates global service delivery.

Four Levels of Federation

Federation is categorized into four distinct hierarchical levels.

- Level 1: Infrastructure Federation (Resource sharing like storage/servers).
- Level 2: Platform Federation (Shared middleware and runtime environments).
- Level 3: Application Federation (Interoperability between software applications).
- Level 4: Data Federation (Unified access to distributed data sources).



Federated Services and Applications

Federated services allow users to access resources across multiple domains.

- Identity Federation: Single Sign-On (SSO) across different cloud platforms.
- Resource Federation: Dynamic allocation of VMs from multiple providers.
- Inter-cloud communication: Standardized protocols like OVF for VM images.
- Benefits: Improved reliability, load balancing, and disaster recovery.
- Challenges: Security, latency, and complex governance models.

Future of Federation

The future of cloud federation focuses on automation and standardization.

- Standardization: Adoption of universal APIs to simplify integration.
- AI-Driven Federation: Intelligent resource brokering based on predictive analytics.
- Edge-Cloud Federation: Integrating edge computing nodes into the federated ecosystem.
- Blockchain: Using distributed ledgers for secure and transparent resource negotiation.
- Multi-cloud management: Tools that abstract the underlying complexity of federated clouds.

Exam corner

- **Q1** Explain the architecture of Hadoop and the role of HDFS.
- **Q2** Describe the MapReduce programming model with a suitable example.
- **Q3** Differentiate between Type 1 and Type 2 hypervisors with VirtualBox as an example.
- **Q4** What is Google App Engine? Explain its Standard vs Flexible environments.
- **Q5** List and explain the core components of OpenStack.
- **Q6** Define Cloud Federation and explain its four levels.
- **Q7** Discuss the challenges and future trends in Cloud Federation.

Unit 5 done. You are exam-ready.

Revise the Exam corner once more, then open Unit 1 on theycotes.com. Every AKTU subject, unit by unit, free. Notes by Chirag Jain, Ycotes.

